

RESEARCH

Open Access

The strength of genetic interactions scales weakly with mutational effects

Andrea Velenich* and Jeff Gore

Abstract

Background: Genetic interactions pervade every aspect of biology, from evolutionary theory, where they determine the accessibility of evolutionary paths, to medicine, where they can contribute to complex genetic diseases. Until very recently, studies on epistatic interactions have been based on a handful of mutations, providing at best anecdotal evidence about the frequency and the typical strength of genetic interactions. In this study, we analyze a publicly available dataset that contains the growth rates of over five million double knockout mutants of the yeast *Saccharomyces cerevisiae*.

Results: We discuss a geometric definition of epistasis that reveals a simple and surprisingly weak scaling law for the characteristic strength of genetic interactions as a function of the effects of the mutations being combined. We then utilized this scaling to quantify the roughness of naturally occurring fitness landscapes. Finally, we show how the observed roughness differs from what is predicted by Fisher's geometric model of epistasis, and discuss the consequences for evolutionary dynamics.

Conclusions: Although epistatic interactions between specific genes remain largely unpredictable, the statistical properties of an ensemble of interactions can display conspicuous regularities and be described by simple mathematical laws. By exploiting the amount of data produced by modern high-throughput techniques, it is now possible to thoroughly test the predictions of theoretical models of genetic interactions and to build informed computational models of evolution on realistic fitness landscapes.

Keywords: Epistasis, Evolution, Fitness landscapes, Genetic interactions, Yeast

Background

Genetic interactions [1] have shaped the evolutionary history of life on earth. They have been found to limit the accessibility of evolutionary paths [2], to confine populations to suboptimal evolutionary states and, on larger time scales, to control the rate of speciation [3]. Epistatic interactions can also be relevant to the development of complex human diseases such as diabetes [4]. Complex traits and diseases are determined by a multiplicity of genomic loci [5], whose independent effects and interactions [6] are often necessary to understand the phenotype of interest. Despite the broad implications of epistatic interactions, a quantitative characterization of their typical strength is still lacking. In this

study, we consider growth rate in yeast as an example of a complex trait modulated by genetic interactions.

Previous studies [7-10] on the relation between the growth effects of a mutation and its epistatic interactions have often been based on a handful of mutations, and only in recent years has anecdotal evidence started being replaced by robust statements based on large data sets. Perhaps the most impressive of these datasets is the one made publicly available with the publication of the article entitled 'The genetic landscape of a cell' by Costanzo *et al.* [11]. The genome of the budding yeast *Saccharomyces cerevisiae* includes approximately 6,000 genes, about 1,000 of which are essential. Viable mutants can be constructed by knocking out any of the approximately 5,000 non-essential genes, by reducing the expression of the essential genes, or by partially compromising the functionality of the gene products. The dataset (see Additional file 1, Figure S1) has been compiled with the growth rates of about 5.4 million double knockout

* Correspondence: velenich@mit.edu
Department of Physics, Massachusetts Institute of Technology, 77
Massachusetts Avenue, Cambridge, MA 02139, USA

mutants, a sizable fraction of all possible double knockout mutants in yeast. Supported by the Costanzo *et al.* dataset, we consider the fundamental question of whether mutations with larger effects have stronger genetic interactions.

Results and discussion

An unbiased definition of genetic interactions

A basic approach to study genetic interactions is to consider two mutations with known effects on a quantitative trait, and to measure their combined effect in the double mutant [12]. Given [11,13] the growth rates of a wild type *S. cerevisiae* strain ($g_{00} = 1$) and of two single knockout mutants (g_{01} and g_{10}), the growth rate of the double knockout mutant (g_{11}) is adequately predicted by a multiplicative null model:

$$g_{11}/g_{00} = (g_{01}/g_{00}) (g_{10}/g_{00}).$$

Equivalently, defining 'log growth' as the logarithm of the relative growth rate,

$$G = \log_2 (g/g_{00}),$$

the log growth of the double knockout mutant is predicted by an additive null model (Figure 1a):

$$G_{11} = G_{01} + G_{10}.$$

Epistatic interactions are identified as deviations from the null model, but several non-equivalent alternatives exist for quantifying these deviations [14]. The most common definition of epistasis considers the difference between the measured and the predicted growth rates for the double knockout mutant [11]:

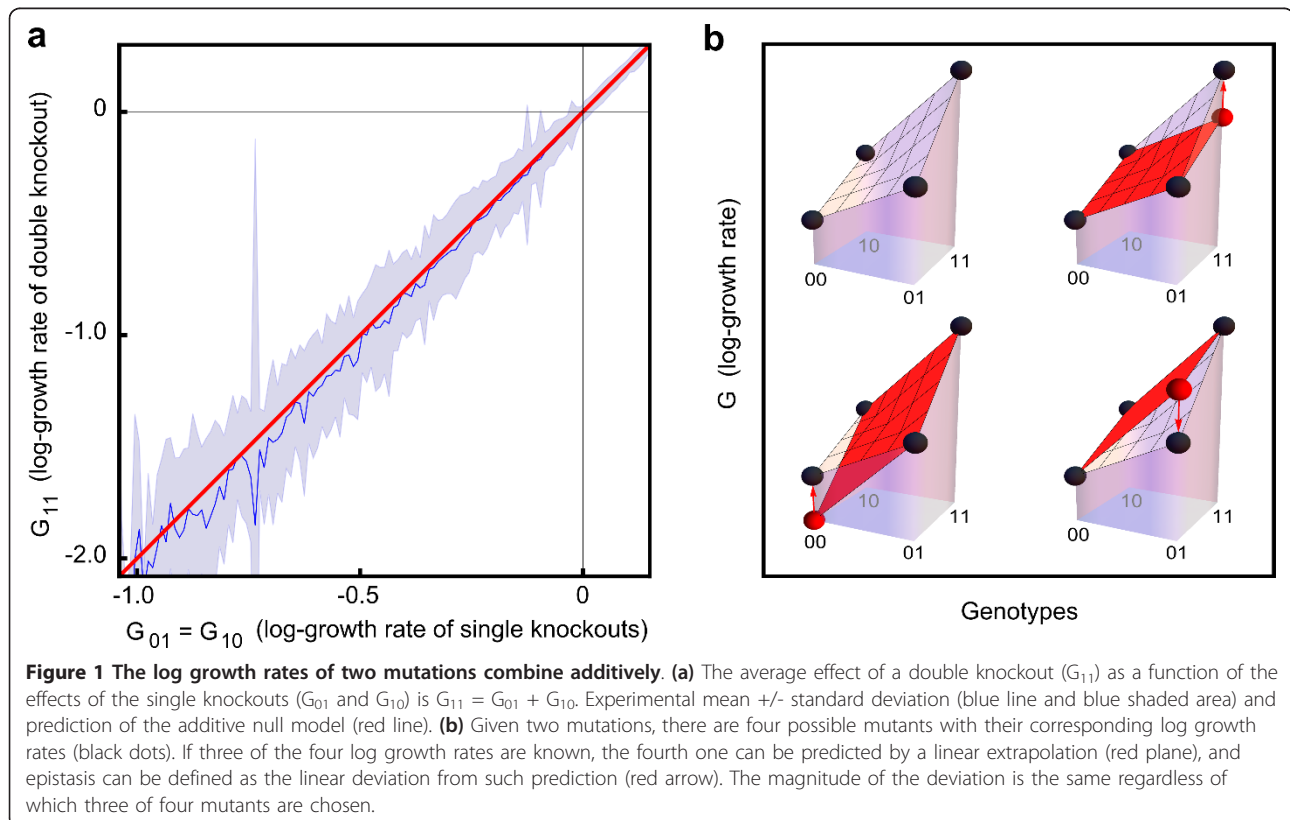
$$e = \frac{g_{11}}{g_{00}} - \frac{g_{01}}{g_{00}} \frac{g_{10}}{g_{00}}$$

Importantly, this definition of e subtly constrains the possible values of epistasis. In fact, when combining very deleterious mutations, e cannot be large and negative even when the double knockout mutant is a synthetic lethal mutant:

$$e = 0 - (g_{01}/g_{00}) (g_{10}/g_{00}) \approx 0, \\ \text{if } g_{01} \ll g_{00} \text{ and } g_{10} \ll g_{00}.$$

In order to avoid *a priori* constraints on the intensity of epistasis, genetic interactions can be defined as the ratio between the measured and predicted relative growth rates, leading to:

$$E = \log_2 \frac{g_{11}}{g_{00}} - \log_2 \frac{g_{01}}{g_{00}} - \log_2 \frac{g_{10}}{g_{00}}.$$



As an example, $E = +1$ indicates a double mutant whose growth rate is twice as large as would be expected based upon the multiplicative null model, whereas $E = -1$ indicates a double mutant whose growth rate is half as large as predicted. This definition of epistasis as fold deviation in the multiplicative model for growth rates is equivalent to a natural definition of epistasis as linear deviation in the additive model for log growth rates (Figure 1b):

$$E = (G_{00} + G_{11}) - (G_{01} + G_{10}) = (G_{11} - G_{01} - G_{10}).$$

A second bias of the common definition of epistasis is that e depends on the choice of which genotype is labeled as 'wild type' or '00', a choice which is always arbitrary, but more obviously so when studying engineered organisms or populations evolving in alternating environments [15]. By contrast,

$$|E| = |(G_{00} + G_{11}) - (G_{01} + G_{10})|$$

depends only on which pair of genes is considered, being a geometric measure for the 'curvature' of the fitness landscape (Figure 1b).

The definition of E has found some favor in the theoretical literature [7,16], but it is not routinely used to analyze experimental data apart from rare exceptions [8,17]. Its main drawback is that synthetic lethals have a log growth rate of $-\infty$, and require a separate although simpler analysis in which lethal interactions can simply be counted. The definition of E proves instead to be extremely valuable when quantifying the strength of non-lethal genetic interactions.

Epistatic interactions scale weakly with mutational effects

With the appropriate definition of epistasis, a simple relation between the growth rate effects of two mutations and the expected strength of their interaction emerges.

Let us consider two groups of mutations; in the first group, all mutations have log growth effect G_{01} , and in the second group, all mutations have log growth effect G_{10} . We can then build all possible double mutants obtained by combining one mutation from each group. In the absence of epistasis, all the double mutants have a log growth rate

$$G_{11} = G_{01} + G_{10},$$

and the distribution of genetic interactions is sharply peaked at $E = 0$. When epistasis is present, the distribution of genetic interactions has, in general, non-zero mean and standard deviation. Experimentally, however, the mean of genetic interactions is close to zero (this is why the null model remains approximately valid) (Figure 1a; Figure 2d). Even when the mean interaction is vanishing, the difference

between the experimental dataset and the ideal case without interactions can be quantified by the finite value of the experimental standard deviation $\sigma(G_{01}, G_{10})$, which provides a numerical estimate for the characteristic strength of epistatic interactions.

In order to produce reliable numerical results, thousands of growth rates are necessary to characterize the probability distribution of epistasis. We analyzed the Costanzo *et al.* dataset by binning pairs of mutations according to the log growth effects of their single knockouts G_{01} and G_{10} , using the method described above to outline the probability distribution of epistasis. We chose bin sizes that grow exponentially with G in order to ensure an approximately constant number of data points in each bin (see Materials and Methods; see Additional file 1, Figure S2). Most bins contain from thousands to tens of thousands of data points. For each bin, we computed

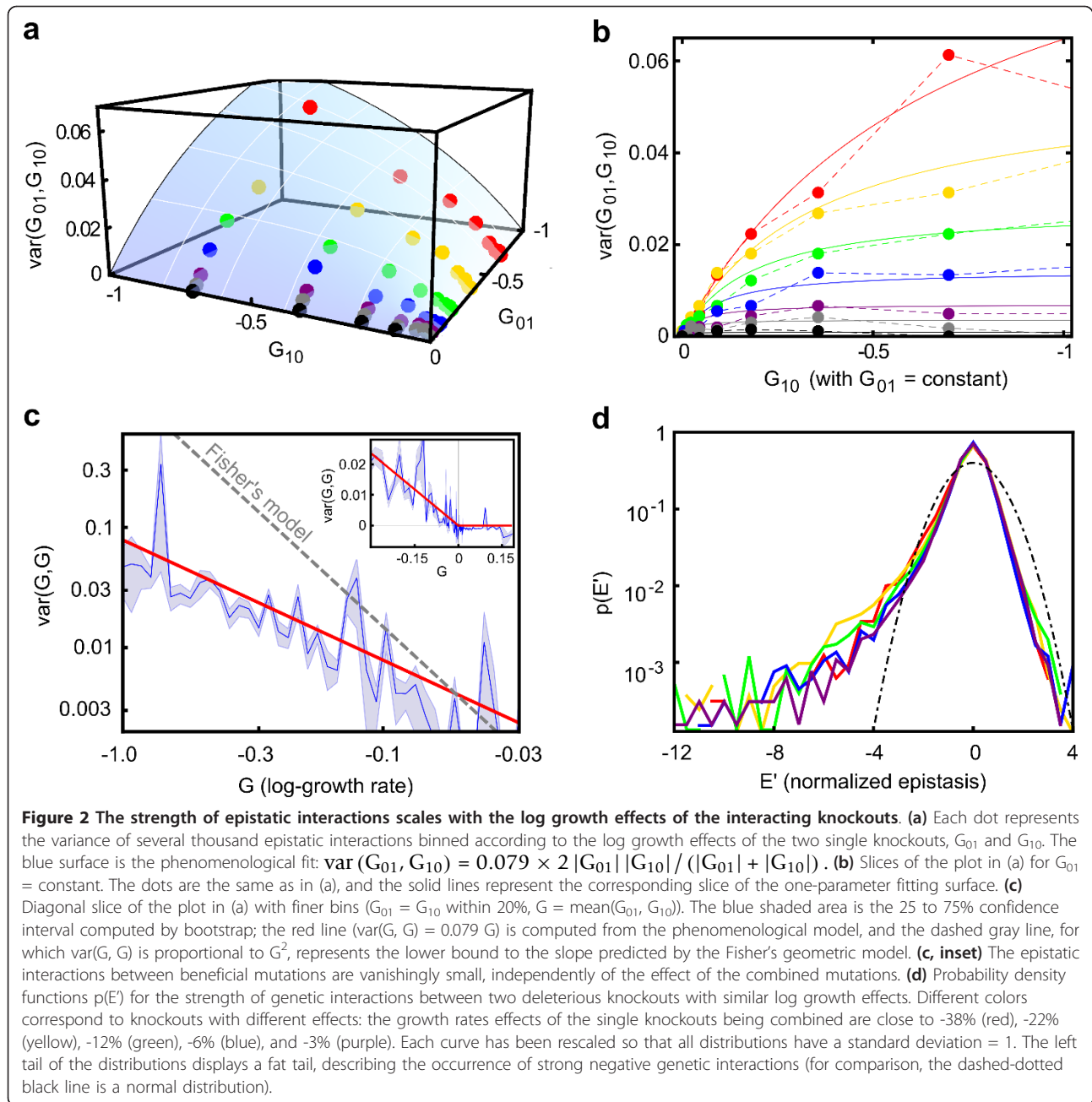
$$\text{var}(E(G_{01}, G_{10})),$$

that is, the variance of the random variable E relative to the bin labeled by growth rates G_{01} and G_{10} . In the rest of the paper we will refer to such variance as $\text{var}(G_{01}, G_{10})$, emphasizing that the variance in the strength of epistatic interactions is, eventually, a function of G_{01} and G_{10} (Figure 2a). The square root of the variance, $\sigma(G_{01}, G_{10})$, then represents the expected strength of epistasis as a function of the independently varying effects of the two single knockouts. A natural expectation for the dependence of epistasis on the effect of the combined mutations comes from rescaling Figure 1a; if all the log growth effects of single and double knockouts increase by a factor of two, then the strength of epistasis should also increase by a factor of two. Unexpectedly, however, when combining deleterious mutations, the strength of epistatic interactions does grow with the effects of the mutations that are combined, but the dependence is much weaker; when the effect of both single knockouts is doubled, the strength of epistasis increases only by a factor of $\sqrt{2}$ (Figure 2).

In more detail, we observed that if the effect of the first knockout (G_{01}) is held constant, the dependence of the variance of epistasis on the effect of the second knockout (G_{10}) is well approximated by a Michaelis-Menten law (Figure 2b):

$$\text{var}(G_{10}) = v \frac{|G_{10}|}{K + |G_{10}|}.$$

When the effects of both knockouts are free to vary, the requirement that the variance is a symmetric function of its two variables, G_{01} and G_{10} , implies that $K = |G_{01}|$ and that v is proportional to G_{01} . A one-parameter



function which fits the seen variance over the whole range of deleterious fitness effects (Figure 2a) is then:

$$\text{var}(G_{01}, G_{10}) = 2c \frac{|G_{01} G_{10}|}{|G_{01}| + |G_{10}|} \text{ with } c = 0.079.$$

This functional form can also be obtained from a simple model based on diffusion in fitness space (see Additional file 1, Supplementary text 1). An even simpler phenomenological fit, although slightly less accurate, is:

$$\text{var}(G_{01}, G_{10}) = c\sqrt{(|G_{01}| |G_{10}|)}$$

(see Additional file 1, Figure S3). Importantly, these functions capture two major features of the data; first, epistasis vanishes when G_{01} or $G_{10} = 0$; second, when the effects of the two knockouts are similar ($G_{01} = G_{10} = G$ along the diagonal of the surface in Figure 2a), the variance of epistasis is approximately proportional to G (Figure 2c):

$$\text{var}(G_{01}, G_{10}) = c|G|.$$

The scaling described above is seen only for deleterious knockouts. When combining the beneficial

knockouts in the dataset instead, the strength of epistasis is close to zero (Figure 2c, inset). This might be because the slightly beneficial knockouts are not adaptive mutations, but simply remove genes that are not needed in the conditions chosen for the experiment, so that their interactions are likely to be negligible. However, in apparent contrast to this observation, recent studies [8,18] on adaptive mutations in *Escherichia coli* suggest that genetic interactions between adaptive mutations are mostly negative. In fact, during adaptation, the prevalence of negative interactions is likely to be caused by biased sampling, because the mutations that fix in the population are likely to be the ones that solve environmental or biological challenges for an organism. Diminishing returns arise because the appearance of multiple 'solutions' to the same challenge is not necessarily preferable over the presence of a single solution. Rather than focusing on mutations that fix during a bout of adaptation, the Costanzo *et al.* dataset includes a large fraction of all possible pairs of genes in the yeast genome. Because for most pairs the two genes are involved in unrelated biological processes, interactions are often vanishingly small. We did observe, however, that the distribution of epistatic interactions is asymmetric, with a heavy tail of deleterious interactions (Figure 2d).

Experimental uncertainty generates spurious epistatic interactions

When inferring genetic interactions from experimental data, it is important to take into account that each measured growth rate is affected by some uncertainty, and that measurement errors in the growth rates could erroneously be interpreted as genetic interactions. Importantly, for each single and double mutant, the Costanzo *et al.* dataset provides the mean growth rate together with its estimated experimental uncertainty (the growth rate of each mutant being measured at least four times).

In order to quantify the effect of the experimental uncertainty on the inferred epistatic interactions, we constructed a number of mock datasets, assuming that the null model without epistatic interactions described biology exactly. In these datasets, each single knockout had the same growth rate as in the original dataset, and each double knockout had a growth rate equal to the product of the relative growth rates of the corresponding single knockouts. We then randomized the mock datasets by shifting each growth rate by a random amount sampled from a Student's *t*-distribution, with width depending on the corresponding experimental uncertainty reported in the original dataset (see Additional file 1, Supplementary text 3). As expected, analysis of these 'noisy' datasets revealed some epistasis, clearly caused by our addition of experimental noise

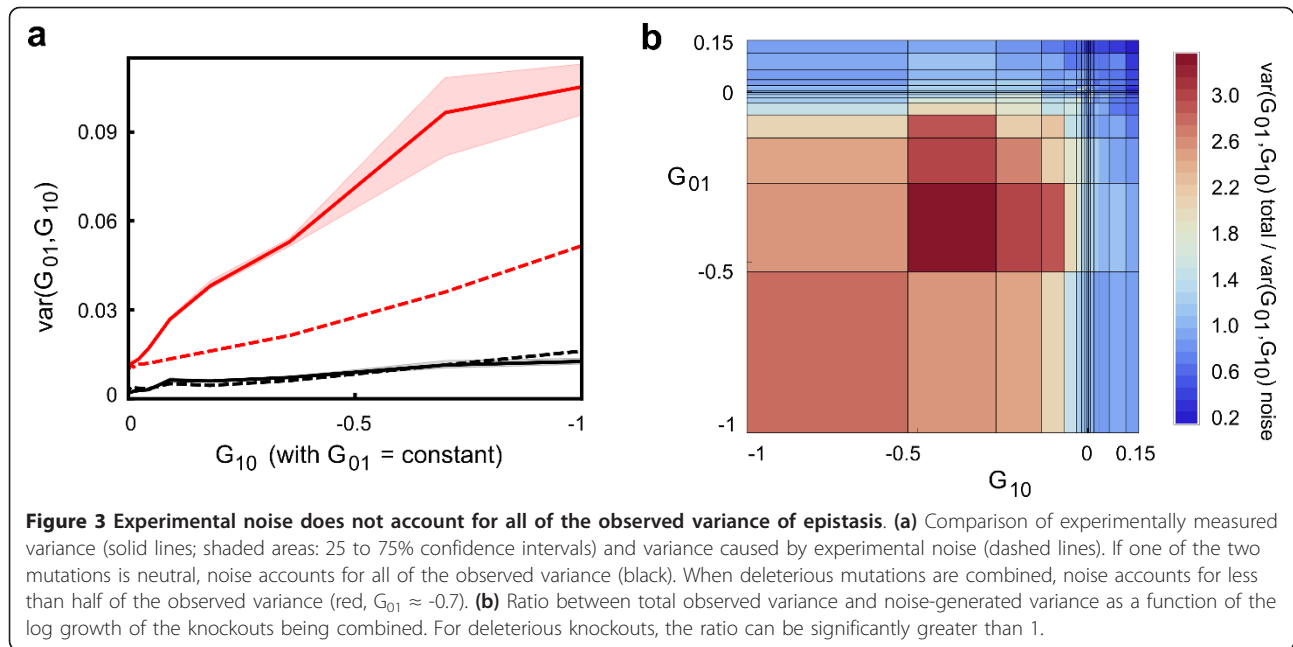
rather than by any biological mechanism. We found that for pairs involving beneficial or neutral mutations, the variance computed in the mock datasets was comparable to or even greater than the variance observed in the original dataset (Figure 3a, black curves; Figure 3b, blue regions). This fact provides an important internal control, suggesting that the experimental noise has not been underestimated. In spite of this, for pairs of knockouts with substantially deleterious effects, experimental noise accounted for less than half of the total observed variance, with the rest representing genuine biological interactions (Figure 3a, red curves; Figure 3b, red regions).

We then decomposed the variance observed in the original dataset into a contribution produced by experimental uncertainty and a contribution of biological origin; the strength of epistatic interactions was finally computed as the square root of the biological part of the variance. For deleterious knockouts, the relative difference between epistasis computed from the raw data and from the data after subtracting the experimental noise was less than 30%, emphasizing the significant but not overwhelming contribution of experimental noise to the observed variability. Figure 2(a-c) represents the 'biological' part of the observed epistasis; before subtracting the contribution of the experimental uncertainty, the plots are qualitatively similar, but quantitatively slightly different (see Additional file 1, Figure S4). Importantly, because variances are additive, the estimated contribution of the experimental uncertainty to epistasis is largely independent of the choice of the statistical distribution used to model experimental uncertainty. In two instances, however, the unknown details of the full distribution of experimental noise are important; when outlining the distribution of epistatic interactions (Figure 2d) and when describing the probability to observe sign epistasis (Figure 4b). In those two figures, we plotted the raw data, and did not attempt to deconvolve the contribution of experimental uncertainty.

Comparison between theory and experiment

The scaling of epistasis observed in the Costanzo *et al.* dataset (Figure 2) is in sharp contrast to the predictions of Fisher's geometric model [19], a popular model of epistasis in which genetic interactions emerge from geometry. As we saw, when the effects of the two knockouts are similar ($G_{01} = G_{10} = G$), the variance of epistasis is approximately proportional to G . By contrast, in the Fisher's model, the variance $\text{var}(G, G)$ would grow faster than G^2 (Figure 2c; see Additional file 1, Supplementary text 2), a much stronger dependence than the linear dependence observed experimentally.

A concrete numerical example can highlight the importance of the weaker-than-expected scaling of epistasis described in this study. Let us consider two gene



knockouts, each of which reduces the relative growth rate by 5%, from 1.0 to 0.95. According to the multiplicative null model, the growth rate of the double knockout will be approximately 0.95^2 , or approximately 0.90. The questions now are: What kind of deviations could be expected around 0.90? Would a growth rate of 0.85 be surprising? What about a growth rate of 0.50? Let us use the analytic fit discussed in the previous section

$$g_{01} = g_{10} = 0.95,$$

Then

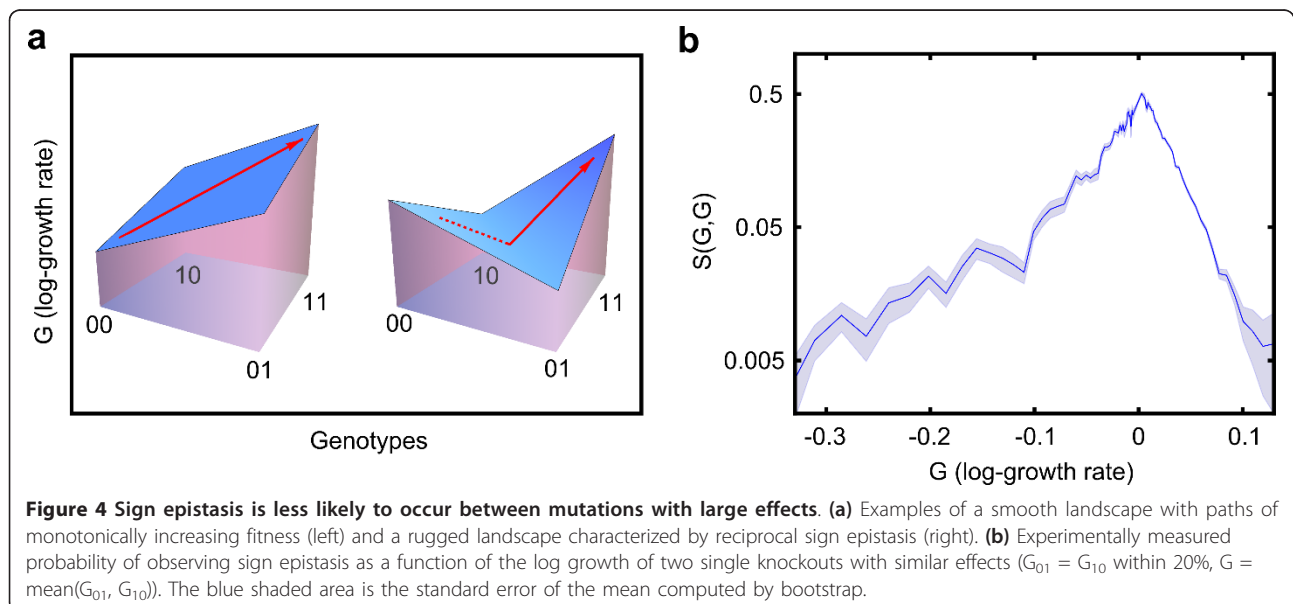
$$G_{01} = G_{10} = \log_2(0.95) = -0.074,$$

$$G_{11} = G_{01} + G_{10} = -0.148,$$

and

$$\sigma(G_{01}, G_{10}) = 0.076.$$

A +/- one standard deviation interval for the growth rate of the double knockout is then



$$[2^{-0.148-0.076}, 2^{-0.148+0.076}] = [0.86, 0.95].$$

Notice that it is not unlikely that epistasis will cancel the effect of the second mutation, so that the growth rate of the double knockout mutant is greater than 0.95, that is, greater than the growth rate of either of the single knockout mutants.

Let us now consider two gene knockouts with stronger effects, each of which reduces the growth rate from 1.0 to 0.60. Then

$$G_{01} = G_{10} = \log_2(0.60) = -0.737,$$

about 10 times as large as the log growth of the single mutants in the previous example. The Fisher's model would predict a $\sigma(G, G)$ at least 10 times larger than in the previous example ($\sigma(G, G) \geq 0.76$), and an interval of likely growth rates for the double knockout mutants at least as large as

$$[2^{-1.47-0.76}, 2^{-1.47+0.76}] = [0.213, 0.610].$$

Notice how, once again, it is not unlikely that owing to genetic interactions, the growth rate of the double knockout mutant is greater than 0.60, the growth rate of either of the two single knockout mutants. The analytic model derived from the experimental data leads to a strikingly different conclusion:

$$\sigma(G_{01}, G_{10}) = 0.241,$$

and the +/- one standard deviation interval for the growth rate of the double knockout becomes

$$[2^{-1.47-0.241}, 2^{-1.47+0.241}] = [0.305, 0.425].$$

In this case, a deviation from the null model that is greater than three standard deviations would be needed for the double knockout mutant to have a growth rate greater than that of the single knockout (0.60), making the event extremely unlikely.

Epistasis constrains the evolutionary dynamics

The previous section provided two examples of reciprocal sign epistasis, realized when two deleterious mutations produce a double mutant that is fitter than either of the two single mutants (Figure 4a). In those cases, a fitness valley limits the evolutionary accessibility of the fitter double mutant, and only on longer time scales may the simultaneous appearance of two mutations [20,21] drive a population to the new local fitness maximum. In this context, the scaling behavior of epistasis is of great importance, because it determines the number and the topology of the evolutionarily accessible paths [2,22,23], ultimately affecting the possible outcomes of the evolutionary process.

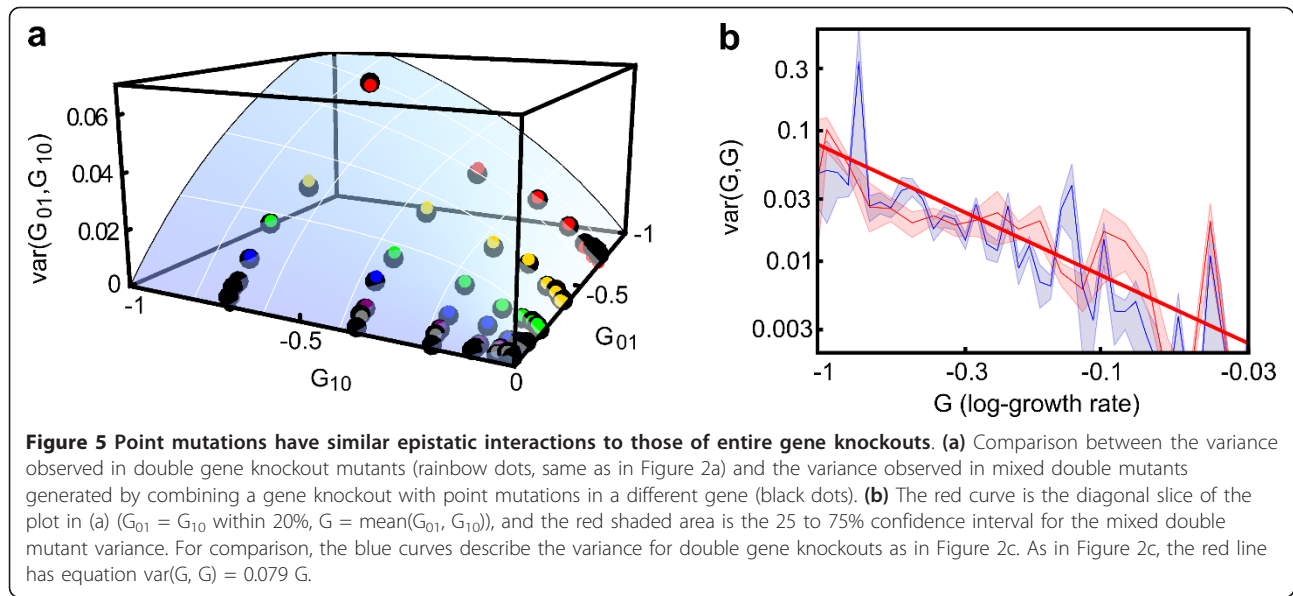
In order to describe how epistasis shapes the naturally occurring fitness landscapes, let us consider $S(G, G)$, the probability to observe sign epistasis when combining two mutations with similar growth rate effects, G . Here, $S(G, G)$ depends on the typical interaction strength,

$$\sigma(G, G) = \sqrt{\text{var}(G, G)}.$$

In particular, if $\sigma(G, G)$ is proportional to G , then the probability of observing sign epistasis is independent of G . The Fisher's model implies a super-linear dependence of $\sigma(G, G)$ on G , thus predicting a greater probability of observing sign epistasis among mutations with strong effects. Instead, if the scaling of $\sigma(G, G)$ is proportional to \sqrt{G} (Figure 2), then sign epistasis is more likely to occur among mutations with small effects (Figure 3b). When the relative growth rate effects of the single knockouts are small (<2 to 3%), experimental uncertainty prevents us from pinpointing which pairs of genes are epistatic. This does not mean, however, that mutations with small effects do not interact. Assuming that the scaling of epistasis we measured directly for mutations with intermediate and large effects extends to mutations with small effects, a consequence of the observed scaling of epistasis is the roughening of the local fitness landscape in the proximity of an evolutionary optimum; when the fitness effects of available mutations become small [24], epistatic interactions become increasingly relevant [25,26], reducing the accessibility of evolutionary paths and further slowing down the rate of adaptation [27,28]. The evolutionary dynamics on correlated fitness landscapes [10,29] with the realistic correlations described here certainly deserves further experimental and theoretical investigation.

The scaling of genetic interactions may be generic

To date, our analysis has been limited to interactions between entire gene knockouts. Although mutations with extreme effects on gene regulation and horizontal gene transfer are biologically relevant mechanisms for the removal or acquisition of whole genes at once, organisms explore possible genetic variants largely through the accumulation of single point mutations. The Costanzo *et al.* data set contains thousands of double mutants for which the first mutation is a gene knockout and the second mutation consists of one or more point mutations in a different gene, causing the gene product to misfold in a temperature-sensitive way. Although the distribution of growth rate effects for point mutations is different than for single gene knockouts (see Additional file 1, Figure S2), the statistics of genetic interactions are remarkably similar when combining two single knockouts and when combining a single knockout with a point mutation (Figure 5). A similar



scaling is also seen for the epistatic interactions between single gene knockouts and decreased abundance by mRNA perturbation [30] (DamP) perturbations of a second gene (see Additional file 1 Figure S5). The analysis of these hybrid double mutants suggests that the statistics of the interactions between any two genetic perturbations are determined only by their growth rate effects [31], and not by their biological origin in terms of point mutations or gene knockouts.

A comparison between different definitions of epistasis

Importantly, any quantitative result on epistasis is a consequence of how epistasis is defined. Of particular interest is how strong an epistatic interaction is deemed to be, based upon its ranking when compared with that of other pairs of mutations. Although the 'traditional' definition

$$e = g_{11}/g_{00} - (g_{01}/g_{00})(g_{10}/g_{00})$$

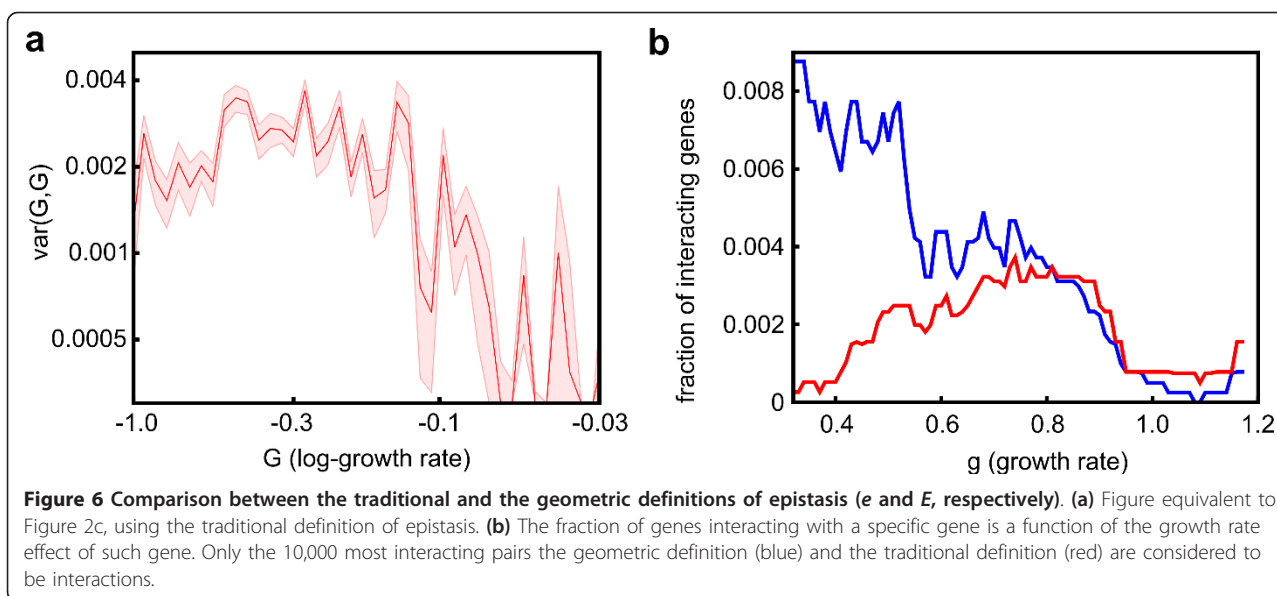
and the 'geometric' definition

$$E = G_{11} - (G_{01} + G_{10})$$

agree about the sets of positive and negative interactions, they assign different strengths and, more importantly, different rankings to the same pair of interacting mutations. As an example, if the Costanzo *et al.* dataset is analyzed using the 'traditional' definition of genetic interactions, then the linear dependence of $\text{var}(G, G)$ on G in Figure 2c is replaced by an oddly non-monotonic dependence, displaying weaker interactions for pairs of genes with either very small or very large fitness effects (Figure 6a). As mentioned previously, this decrease in the inferred strength of epistatic interactions for very

deleterious mutations is a mathematical consequence of the traditional definition of epistasis, rather than a property of genetic interactions. The same bias would lead us to conclude that genes with strong effects on growth are almost non-interacting (Figure 6b, red line). However, because previous studies have determined that essential genes partake in more interactions than do non-essential genes [32], it is also reasonable to expect that non-lethal genes with large growth effects are involved in more interactions than genes with small growth effects. Indeed, according to the 'geometric' definition of epistasis, the fraction of genes with which a gene interacts steadily increases with the growth rate effect of the gene (Figure 6b, blue line). By contrast, the traditional definition of epistasis, consistently assigns low rankings to interactions between genes with large growth rate defects, as confirmed by a further analysis comparing the two definitions of epistasis against interactions inferred from the Gene Ontology (GO) database [33] (see Additional file 1, Figure S6). According to the geometric definition of epistasis, genetic networks [34] are denser than expected not only among essential gene [32], but also among genes with large growth effects.

Finally, it is important to emphasize that the traditional definition of epistasis remains slightly more successful at discovering the functional relations between genes, as cataloged in the GO database (see Additional file 1, Figure S6). Part of the reason for this could be that some of those functional characterizations were suggested by the traditional definition of epistasis in the first place. It is certainly true, however, that many of the top-ranking interactions according to the geometric definition of epistasis involve single and double mutants with small growth rates; for those mutants, experimental



noise is relatively large, and this may cause a few weakly interacting pairs to be incorrectly ranked as strongly interacting. It is likely that the experimental protocols could be easily adjusted to reduce the relative uncertainty on the growth rate of especially slow-growing mutants to avoid this issue (for example, by allowing for a much longer time for growth or by measuring the growth rates of additional replicates).

Conclusions

We analyzed the growth rates of about five million double mutants in the dataset associated with the work by Costanzo *et al.* We characterized how the strength of genetic interactions depends on the growth effects of the mutations being combined, and found a weaker dependence than that predicted by current theoretical models. Although the results were obtained mainly from entire gene knockouts, there is some evidence that the observed scaling might extend to the interactions between single point mutations. The scaling of epistasis might or might not be generic [35,36]; important drivers could be the harshness of the environment [37], details about the evolutionary history [38-40], sexual versus asexual reproduction [41] and, perhaps most importantly, metabolic [42-45] and genetic complexity [46,47]. In general, the experimentally observed scaling suggests a previously unexplored class of correlated fitness landscapes with tunable roughness, in which epistasis depends explicitly on the effects of the mutations being combined.

A clear limitation of our discussion is that only pair interactions were considered. Although high-throughput experiments will provide data on higher-order interactions, a solid understanding of pair interactions remains

necessary before addressing n -mutation interactions. A genuine three-mutation interaction, for instance, should be defined as the unexplained deviation from what can be computed by combining the effects of all relevant mutations and their pair interactions [10,48], perhaps using linear fits within the additive null model for log growth rates.

The results we present here were based on a geometric definition of epistasis. We compared this definition with a more standard definition, highlighting the desirable mathematical properties of the geometric definition and the simple phenomenological relations it produces.

In conclusion, although each epistatic interaction between specific genes depends on biological details and remains largely unpredictable from first principles, we have shown that the statistical properties of an ensemble of interactions can display conspicuous regularities, and can be described by simple mathematical laws.

Materials and methods

The Costanzo *et al.* dataset is publicly available [49]. The file `sgadata_costanzo2009_rawdata_101120.txt.gz` was downloaded on August 17, 2010 and analyzed with Mathematica (code available at the Gore laboratory website [50]). We restricted our analysis to double knockout mutants whose growth rates were positive numerical values and for which the growth rates of both single mutants were numerical values (see Additional file 1, Figure S1). Some genes appear in the dataset both as query and array genes; care was taken to avoid double counting.

The exponentially growing intervals used for the binning of the log growth rate effects were defined as $[-2^n, -2^{n-1}]$

for an appropriate range of integer n 's. Owing to the rarity of extremely deleterious mutations, bins for positive n 's contained only a few data points, while bins with large negative n 's were extremely small. In the figures we reported only bins for $n = -7$ to 0 , containing log growth rate effects ranging from $-2^0 = -1$ to $-2^{-8} = -0.0039$ or, alternatively, relative growth rate effects ranging from $2^{-1} = 0.5$ to $2^{-0.0039} = 0.997$. Different choices for the binning sizes and positions did not significantly alter the results of the analysis.

In order to quantify the contribution of experimental uncertainty to epistasis, we generated nine randomized mock datasets. The mean level of noise-generated epistasis in these nine datasets is reported in Figure 4 (dashed lines), and we provide an extensive discussion of the choice of Student's t -distributions to generate the mock datasets from the original dataset (see Additional file 1, Supplementary text 3).

The GO database `go_201207-assocdb-tables.tar.gz` was downloaded from the GO site [51] on July 19, 2012. The MySQL database was queried with Python and analyzed Mathematica (code available upon request).

Additional material

Additional file 1: Supplementary Figures and Text.

Abbreviations

DamP: Decreased abundance by mRNA perturbation; GO: Gene Ontology.

Conflict of interest

The authors declare that they have no conflict of interest.

Authors' contributions

AV and JG designed research; AV performed research and analyzed data; AV and JG wrote the paper. Both authors have read and approved the final manuscript.

Acknowledgements

We are grateful to Mingjie Dai for collaboration during the early stages of the study. We thank Kirill Korolev, Pankaj Mehta, and the members of the Gore laboratory for providing comments and advice on the manuscript. This research was funded by an NIH Pathways to Independence Award, NSF CAREER Award, Pew Biomedical Scholars Program, and Alfred P. Sloan Foundation Fellowship.

Received: 11 February 2013 Revised: 1 July 2013

Accepted: 26 July 2013 Published: 26 July 2013

References

- Phillips PC: Epistasis - The essential role of gene interactions in the structure and evolution of genetic systems. *Nat Rev Genet* 2008, **9**:855-867.
- Weinreich DM, Delaney NF, DePristo MA, Hartl DL: Darwinian evolution can follow only very few mutational paths to fitter proteins. *Science* 2006, **312**:111-114.
- Dettman JR, Sirjusingh C, Kohn LM, Anderson JB: Incipient speciation by divergent adaptation and antagonistic epistasis in yeast. *Nature* 2007, **447**:585-588.
- Hoh J, Ott J: Mathematical multi-locus approaches to localizing complex human trait genes. *Nat Rev Genet* 2003, **4**:701-709.
- Mackay TFC, Stone EA, Ayroles JF: The genetics of quantitative traits: challenges and prospects. *Nat Rev Genet* 2009, **10**:565-577.
- Jansen RC: Studying complex biological systems using multifactorial perturbation. *Nat Rev Genet* 2003, **4**:145-151.
- Gros PA, Le Nagard H, Tenaillon O: The evolution of epistasis and its links with genetic robustness, complexity and drift in a phenotypic model of adaptation. *Genetics* 2009, **182**:277-293.
- Khan AI, Dinh DM, Schneider D, Lenski RE, Cooper TF: Negative epistasis between mutations in an evolving bacterial population. *Science* 2011, **332**:1193-1196.
- Wilke CO, Adami C: Interaction between directional epistasis and average mutational effects. *Proc R Soc Lond B Biol Sci* 2001, **268**:1469-1474.
- Beerenwinkel N, Pachter L, Sturmfeils B, Elena SF, Lenski R: Analysis of epistatic interactions and fitness landscapes using a new geometric approach. *BMC Evol Biol* 2007, **7**:60-73.
- Costanzo M, Baryshnikova A, Bellay J, Kim Y, Spear ED, Sevier CS, Ding H, Koh JLY, Toufighi K, Mostafavi S, Prinz J, St Onge RP, VanderSluis B, Makhnevych T, Vizeacoumar FJ, Alizadeh S, Bahr S, Brost RL, Chen Y, Cokol M, Deshpande R, Li Z, Lin Z-Y, Liang W, Marback M, Paw J, San Luis B-J, Shuteriqi E, Tong AHY, van Dyk N, et al: The genetic landscape of a cell. *Science* 2010, **327**:425-431.
- Dixon SJ, Costanzo M, Baryshnikova A, Andrews B, Boone C: Systematic mapping of genetic interaction networks. *Annu Rev Genet* 2009, **43**:601-625.
- Baryshnikova A, Costanzo M, Kim Y, Ding H, Koh J, Toufighi K, Youn J-Y, Ou J, San Luis B-J, Bandyopadhyay S, Hibbs M, Hess D, Gingras A-C, Bader GD, Troyanskaya OG, Brown GW, Andrews B, Boone C, Myers CL: Quantitative analysis of fitness and genetic interactions in yeast on a genome scale. *Nat Methods* 2010, **7**:1017-1024.
- Mani R, StOnge RP, Hartman JL, Giaever G, Roth FP: Defining genetic interaction. *Proc Natl Acad Sci USA* 2008, **105**:3461-3466.
- Tan L, Gore J: Slowly switching between environments facilitates reverse evolution in small populations. *Evolution* 2012, **66**:3144-3154.
- Peters AD, Lively CM: Epistasis and the maintenance of sex. In *Epistasis and the evolutionary process*. Edited by: Wolf JB, Brodie ED, Wade MJ. Oxford: Oxford University Press; 2000:99-112.
- Martin G, Elena SF, Lenormand T: Distributions of epistasis in microbes fit predictions from a fitness landscape model. *Nat Genet* 2007, **39**:555-560.
- Chou HH, Chiu HC, Delaney NF, Segre D, Marx CJ: Diminishing return epistasis among beneficial mutations decelerates adaptation. *Science* 2011, **332**:1190-1192.
- Fisher RA: *The Genetical Theory of Natural Selection*. Oxford: Clarendon Press; 1930.
- Weinreich DM, Watson RA, Chao L: Sign epistasis and genetic constraint on evolutionary trajectories. *Evolution* 2005, **59**:1165-1174.
- Weissman DB, Desai MM, Fisher DS, Feldman MW: The rate at which asexual populations cross fitness valleys. *Theor Pop Biol* 2009, **75**:286-300.
- Poelwijk FJ, Kiviet DJ, Weinreich DM, Tans ST: Empirical fitness landscapes reveal accessible evolutionary paths. *Nature* 2007, **445**:383-386.
- Velenich A, Gore J: Synthetic approaches to understanding biological constraints. *Curr Opin Chem Biol* 2012, **16**:323-328.
- Eyre-Walker A, Keightley PD: The distribution of fitness effects of new mutations. *Nat Rev Genet* 2007, **8**:610-618.
- Tan L, Serene S, Chao HX, Gore J: Hidden randomness between fitness landscapes limits reverse evolution. *Phys Rev Lett* 2011, **106**:198102.
- Woods RJ, Barrick JE, Cooper TF, Shrestha U, Kauth MR, Lenski RE: Second-order selection for evolvability in a large *Escherichia coli* population. *Science* 2011, **331**:1433-1436.
- Orr HA: The population genetics of adaptation the distribution of factors fixed during adaptive evolution. *Evolution* 1998, **52**:935-949.
- Orr HA: The genetic theory of adaptation a brief history. *Nat Rev Genet* 2005, **6**:119-127.
- Kryazhimskiy S, Tkčik G, Plotkin JB: The dynamics of adaptation on correlated fitness landscapes. *Proc Natl Acad Sci USA* 2009, **106**:18638-18643.
- Breslow DK, Cameron DM, Collins SR, Schuldiner M, Stewart-Ornstein J, Newman HW, Braun S, Madhani HD, Krogan NJ, Weissman JS: A comprehensive strategy enabling high-resolution functional analysis of the yeast genome. *Nat Methods* 2008, **5**:711-718.
- Xu L, Barker B, Gu Z: Dynamic epistasis for different alleles of the same gene. *Proc Natl Acad Sci USA* 2012, **109**:10420-10425.

32. Davierwala AP, Haynes J, Li Z, Brost RL, Robinson MD, Yu L, Mnaimneh S, Ding H, Zhu H, Chen Y, Cheng X, Brown GW, Boone C, Andrews BJ, Hughes TR: **The synthetic genetic interaction spectrum of essential genes.** *Nat Genet* 2005, **37**:1147-1152.
33. Barabási AL, Oltvai ZN: **Network biology: understanding the cell's functional organization.** *Nat Rev Genet* 2004, **5**:101-113.
34. The Gene Ontology Consortium: **Gene ontology: tool for the unification of biology.** *Nat Genet* 2000, **25**:25-29.
35. Dixon SJ, Fedysyn Y, Koh JLY, Keshava Prasad TS, Chahwan C, Chua G, Toufighi K, Baryshnikova A, Hayles J, Hoe K-L, Kim D-U, Park H-O, Myers CL, Pandey A, Durocher D, Andrews BJ, Boone C: **Significant conservation of synthetic lethal genetic interaction networks between distantly related eukaryotes.** *Proc Natl Acad Sci USA* 2008, **105**:16653-16658.
36. Tischler J, Lehner B, Fraser AG: **Evolutionary plasticity of genetic interaction networks.** *Nat Genet* 2008, **40**:390-391.
37. Harrison R, Papp B, Pál C, Oliver SG, Delneri D: **Plasticity of genetic interactions in metabolic networks of yeast.** *Proc Natl Acad Sci USA* 2007, **104**:2307-2312.
38. Wagner A: **Gene duplications robustness and evolutionary innovations.** *BioEssays* 2008, **30**:367-373.
39. Wagner A: **Distributed robustness versus redundancy as causes of mutational robustness.** *BioEssays* 2005, **27**:176-188.
40. Roguev A, Bandyopadhyay S, Zofall M, Zhang K, Fischer T, Collins SR, Qu H, Shales M, Park H-O, Hayles J, Hoe K-L, Kim D-U, Ideker T, Grewal SI, Weissman JS, Krogan NJ: **Conservation and rewiring of functional modules revealed by an epistasis map in fission yeast.** *Science* 2008, **322**:405-410.
41. Azevedo RBR, Lohaus R, Srinivasan S, Dang KK, Burch CL: **Sexual reproduction selects for robustness and negative epistasis in artificial gene networks.** *Nature* 2006, **440**:87-90.
42. Segrè D, DeLuna A, Church GM, Kishony R: **Modular epistasis in yeast metabolism.** *Nat Genet* 2005, **37**:77-83.
43. He X, Qian W, Wang Z, Li Y, Zhang J: **Prevalent positive epistasis in *Escherichia coli* and *Saccharomyces cerevisiae* metabolic networks.** *Nat Genet* 2010, **42**:272-276.
44. Szappanos B, Kovács K, Szamecz B, Honti F, Costanzo M, Baryshnikova A, Gelius-Dietrich G, Lercher MJ, Jelasity M, Myers CL, Andrews BJ, Boone C, Oliver SG, Pál C, Papp B: **An integrated approach to characterize genetic interaction networks in yeast metabolism.** *Nat Genet* 2011, **43**:656-662.
45. Almaas E, Kovács B, Vicsek T, Oltvai ZN, Barabási AL: **Global organization of metabolic fluxes in the bacterium *Escherichia coli*.** *Nature* 2004, **427**:839-843.
46. Sanjuán R, Elena SF: **Epistasis correlates to genomic complexity.** *Proc Natl Acad Sci USA* 2006, **103**:14402-14405.
47. Sanjuán R, Nebot MR: **A network model for the correlation between epistasis and genomic complexity.** *PLoS One* 2008, **3**:e2663.
48. Wood K, Nishida S, Sontag ED, Cluzel P: **Mechanism-independent method for predicting response to multidrug combinations in bacteria.** *Proc Natl Acad Sci USA* 2012, **109**:12254-12259.
49. **The Genetic Landscape of the Cell.** [http://drygin.ccbbr.utoronto.ca/~costanzo2009].
50. **Gore Laboratory.** [http://www.gorelab.org/software.html].
51. **Gene Ontology. GO Database Downloads.** [http://www.geneontology.org/GO.downloads.database.shtml].

doi:10.1186/gb-2013-14-7-r76

Cite this article as: Velenich and Gore: **The strength of genetic interactions scales weakly with mutational effects.** *Genome Biology* 2013 **14**:R76.

Submit your next manuscript to BioMed Central and take full advantage of:

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at
www.biomedcentral.com/submit

